

Remarks

The Present Invention and the Pending Claims

The present invention relates generally to the field of speech recognition. More particularly, the invention discloses a technique for disambiguating speech input using one of voice mode interaction, visual mode interaction, or a combination of voice mode and visual mode interaction.

Claims 1, 4-5, 7-8, and 11-14 are currently pending. Reconsideration and allowance of the pending claims is respectfully requested.

Summary of the Office Action

Claims 1, 4, 7-8, 11 and 14 rejected under 35 U.S.C. 103(a) by **Lai et al. (USPN 6,006,183)** referred to as **Lai** hereinafter in view of **Duan et al. (US Patent No. 6,223,150)** referred to as **Duan** hereinafter.

Claims 5 and 12-13 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Lai** in view of **Haddock et al. (USPN 5,265,014)** referred to as **Haddock** hereinafter.

Amendments To The Claims

Claims 1 and 11 are currently amended. Support for the amendments in claims 1 and 11 are found at paragraphs [0017], [0021], [0022], [0024], [0025], [0027], and [0028].

The office action states: “Claims 1, 4, 7-8, 11 and 14 rejected under 35 U.S.C. 103(a) by **Lai et al. (USPN 6,006,183)** referred to as **Lai** hereinafter in view of **Duan et al. (US Patent No. 6,223,150)** referred to as **Duan** hereinafter.”

First, Lai in view of Duan, does not teach or suggest all the claim limitations. Applicant’s discloses an options and parameters **114** component for setting parameters

required for controlling the disambiguation mechanism by **the user and the application** (see paragraphs [0017], [0021], [0022], [0024], [0025], and Fig 1; paragraph [0024] recites: “*The end user 108 and the application 106 can both set parameters 114 to control the sub-components of the MDM*”). In contrast to the setting of parameters by both the user and the application to control the speech disambiguation mechanism in applicant’s invention, neither Lai nor Duan teach or suggest setting of parameters by the application. Lai teaches assigning a confidence level score by a confidence level scorer **200** of the speech engine **160** (Lai, col. 3, lines 29-30); enabling a user of the system to select score thresholds (Lai, col. 3, lines 37-42); and allowing the user application to accept information from the user control (Lai, col. 4, lines 11-15). Accordingly, Lai in view of Duan does not teach or suggest the following limitation in claim 1:

“an options and parameters component for receiving user parameters and application parameters for controlling the speech disambiguation mechanism, wherein the speech disambiguation mechanism is controlled by parameters set by the user and parameters set by the application”

The Office Action also states: “Lai fails to specifically disclose a plurality of alternatives are presented to a user for selection. In an analogous art Duan discloses providing a user with at least two possible tokens or word disambiguation alternatives... Duan, col. 17, lines 19-36; Figs 13-15”.

In response, in applicant’s disclosure, the speech is disambiguated without translation of the speech input, i.e. the speech disambiguation process is based on the presentation of a single language to the automatic speech recognition system. In contrast, the speech disambiguation process in Duan requires translation of the speech input, and is a two step process. In the first step, alternative words are presented to the user for selection of a disambiguated word. In the second step, the disambiguated word selected in the first step by the user is translated from the source language to the target language. Accordingly, Lai in view of Duan does not teach or suggest the following limitation in claim 1:

“an output interface that presents the selected alternative without translation of the speech input to the application as input.”

Furthermore, applicant discloses presentation of the alternatives to the user in one of voice mode, visual mode, or a **combination of voice and visual mode**, and receiving a selection of an alternative from the user in one of voice mode, visual mode, or a combination of voice mode and visual mode (see paragraph [0010] and paragraph [0028]). In contrast, neither Lai or Duan teach or suggest presenting alternatives to the user in a “a combination of voice and visual mode”, and “receiving a selection of an alternative by the user from the plurality of alternatives presented to the user in one of the voice mode, the visual mode, or a combination of the voice mode and visual mode;”. Accordingly, Lai in view of Duan does not teach the following limitation in claim 1:

“one or more disambiguation components that present the alternatives to the user in one of voice mode, visual mode, or a combination of the voice mode and the visual mode, and receive receives an alternative selected by the user in one of the voice mode, the visual mode, or a combination of the voice mode and the visual mode”

In summary, Lai in view of Duan does not teach the following limitations in claim 1 and 11:

“an options and parameters component for receiving user parameters and application parameters for controlling the speech disambiguation mechanism, wherein the speech disambiguation mechanism is controlled by parameters set by the user and parameters set by the application, …” in claim 1,

“an output interface that presents the selected alternative without translation of the speech input to the application as input” in claim 1,

“one or more disambiguation components that present the alternatives to the user in one of voice mode, visual mode, or a combination of the voice mode and the visual mode, and receive receives an alternative selected by the user in one of the voice mode, the visual mode, or a combination of the voice mode and the visual mode” in claim 1,

“receiving user parameters and application parameters for controlling the speech disambiguation mechanism, wherein both the user and the application can set the parameters to control said speech disambiguation mechanism, and wherein the parameters include confidence thresholds governing unambiguous recognition and close matches” in claim 11,

“presenting the alternatives to the user in one of voice mode, visual mode, or a combination of the voice mode and the visual mode, and receiving a selection of an alternative from the user from the plurality of alternatives presented to the user in one of the voice mode, the visual mode, or a combination of the voice mode and the visual mode” in claim 11, and,

“communicating the selected alternative without translation of the speech input as input to the application”.

Accordingly, applicant respectfully submits that claim 1 and 11 are not obvious over Lai in view of Duan and the rejection of claim 1 and 11 be withdrawn.

Furthermore, applicant respectfully submits that the Lai and Duan references that are sought to be combined are in non-analogous arts. Applicant's invention is a speech recognition system where the speech is disambiguated based on recognition of the speech uttered by the user in a single language by a speech recognition system (see paragraph [0011]). In contrast, Duan is a language translation system where tokens are generated based on translation of a user's utterance from a source language to a target language (see Duan, col. 21, lines 9-13; and col. 9, lines 55-65). A person of ordinary skill in the art confronted with the problem of generating and presenting alternative words to a user for selection of an alternative word in applicant's speech recognition field, will not likely turn to a language translation system to find a solution to the problem. The method by which speech is disambiguated by a speech recognition system without translation of the speech input is vastly different and distinguishable from the method in Duan where the disambiguated word selected in the first step by the user is translated from the source

language to the target language. Therefore, applicant respectfully submits that the teachings of Lai and Duan may not be combined.

Claims 4, 7-8 are dependent on claim 1, and claim 14 is dependent on claim 11. Since Lai and Duan does not teach or suggest the limitations in claim 1 and 11, applicant respectfully submits that claims 4, 7-8, and 14 also not obvious over Lai in view of Duan and the rejection of claims 1, 4, 7-8, 11, and 14 be withdrawn.

The office action also states: “Claims 5 and 12-13 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Lai** in view of **Haddock et al. (USPN 5,265,014)** referred to as **Haddock** hereinafter”.

First, Lai in view of Haddock does not teach all the claim limitations. Applicant discloses disambiguation components that present the alternatives to the user in a visual form and **allow the user to select from among the alternatives** using a voice input (see paragraphs [0018], [0026], and [0027]). Applicant’s process disambiguates the user input **by providing alternatives to each word input** by the user. In contrast, Lai discloses assigning a confidence score to each word by a confidence level scorer and presenting each word/score pair to the user via a Graphical User Interface (GUI).

Even if Lai and Haddock are combined as suggested in the office action, the combination that results will be inoperable for the purpose and functionality of claim 1, i.e., “an options and parameters component for receiving user parameters and application parameters for controlling the speech disambiguation mechanism, wherein the speech disambiguation mechanism is controlled by parameters set by the user and parameters set by the application, ...”. In applicant’s speech disambiguation process, the parameters are set by both the user and the application. In contrast, neither Lai nor Haddock teach or suggest setting of the parameters by both the user and the application.

Applicant’s system for disambiguating speech input comprises, in part, of a disambiguation component that presents two or more alternatives to a user in voice mode,

visual mode, or a **combination of voice mode and visual mode**, and receives an **alternative selected by the user** in voice mode, visual mode, or a **combination of voice mode and visual mode** (see paragraph [0028]). However, neither Lai nor Haddock teach selection of the alternatives by the user using a combination of voice mode and visual mode. As stated on page 2 and 3, items i. to v. of the office action and Figure 2, Lai suggests that an acoustic signal may be inputted to the speech recognizer **190** and the system may display words with confidence level indicated. However, Lai does not teach or suggest that the user can select the disambiguated word in voice mode or visual mode, or a combination of voice mode and visual mode, but instead teaches that the input to the speech recognizer **190** is in voice mode and preferences are provided to the user in visual mode. Hence, there is no teaching or suggestion in Lai and Haddock of the following limitations recited in applicant's claims 5, 12 and 13:

“the disambiguation components present the alternatives to the user in a visual form and allow the user to select from among the alternatives using a voice input” of claim 5,

“the interaction comprises the concurrent use of said visual mode and said voice mode” of claim 12, and

“the interaction comprises the user selecting from among the plural alternatives using a combination of speech and visual-based input” of claim 13.

Furthermore, applicant discloses that the selection algorithm selects the alternatives and presents to the user based on individual confidence values, application parameters, and user parameters (see paragraph [0026]). In contrast, Haddock teaches displaying results to the user based on **previous queries**. Moreover, Haddock does not teach or suggest display of results to the user when the very first query results in ambiguity. Therefore, Lai in view of Haddock does not teach or suggest the following limitation in claim 1:

“a selection component that identifies, according to a selection algorithm, two or more of the tokens to be presented to the user”.

Claim 5 is dependent on claim 1, and claims 12 and 13 are dependent on claim 11. Since claim 1 and 11 are allowable, applicant respectfully submits that claim 5, 12 and 13 are also allowable.

Furthermore, common sense dictates that a person of ordinary skill in the art, at the time the invention was made, would not combine the speech recognition system of Lai and the method for disambiguating natural language queries using referential input by a user as described in Haddock, to arrive at the claimed invention because Lai in view of Haddock show no recognition or appreciation of the following limitations recited in claims 5, 12, and 13:

“the disambiguation components present the alternatives to the user in a visual form and allow the user to select from among the alternatives using a voice input” of claim 5,

“the interaction comprises the concurrent use of said visual mode and said voice mode” of claim 12, and

“the interaction comprises the user selecting from among the plural alternatives using a combination of speech and visual-based input” of claim 13.

Furthermore, in applicant’s disclosure an option is provided to the user for selecting the correct uttered word from a plurality of alternate words, if the speech disambiguation system fails to recognize the correct uttered word. Moreover, the speech disambiguation system enables the user to select the correct word in visual mode, voice mode or a combination of voice and visual mode. Since the applicant’s invention offers wider choice and flexibility to the user for selecting the correct uttered word while using a speech disambiguation system, it is more likely to be commercially successful.

Furthermore, the existing art fails to provide an option to set parameters based on the end application and requires resetting of the parameters each time based on the application. Hence, there is a long felt but unsolved need for an automatic speech recognition system that is initialized based on the user preferences as well as the end application.

In contrast to the above indicia of non-obviousness, Duan, Lai, and Haddock fail to teach or suggest a method of automatic speech recognition with parameter setting based on user preferences and the end application.

For the reasons stated above, applicant respectfully submits that claims 5, 12, and 13 are not obvious over the cited references, and applicant solicits reconsideration of the rejection and allowance of claims 5, 12, and 13.

Conclusion

Applicant respectfully requests that a timely Notice of Allowance be issued in this case. If, in the opinion of Examiner Rider a telephone conference would expedite the prosecution of this application, Examiner Rider is requested to call the undersigned.

Respectfully submitted,

Date: Sep. 24, 2008

Ashok Tankha, Esq.
Attorney For Applicant
Reg. No. 33,802
Phone: 856-266-5145

Correspondence Address

36 Greenleigh Drive
Sewell, NJ 08080
Fax: 856-374-0246